

Redes Genicas: genes esenciales en la red reguladora de *Saccharomyces cerevisiae*

Enciso C., Gutiérrez H.

Departamento de Física

Universidad Nacional de Colombia

Octubre, 2024

Abstract

Este estudio analiza la red reguladora de *Saccharomyces cerevisiae* utilizando dinámica de grafos y medidas de centralidad para identificar factores esenciales en la regulación génica en función de los roles que cada agente desempeña dentro de la jerarquía de la red. Se encontró que la distribución de grados de la red sigue una ley de potencia con un exponente de $\alpha = 2.6347$, sugiriendo una estructura "libre de escala" donde ciertos nodos clave o hubs controlan múltiples conexiones, mientras que la mayoría de los nodos presentan pocos enlaces.

Utilizando el paquete NetworkX en Python, se evaluaron diversas centralidades, revelando que genes como *glc[e]* (glucosa extracelular) y el factor de transcripción GCN4 tienen alta centralidad de grado de salida, lo que indica su relevancia en el metabolismo celular y la respuesta a cambios ambientales. La centralidad de entrada, en cambio, muestra una distribución similar a una de Poisson, lo que sugiere que pocos factores regulan la mayoría de los genes objetivo.

Los nodos con alta proximidad, como GCN4, son esenciales para conectar rutas regulatorias y afectan rápidamente múltiples procesos metabólicos. El estudio concluye que, aunque la red es robusta ante la eliminación aleatoria de nodos, la pérdida de un hub crítico podría fragmentarla, subrayando la importancia de estos nodos en la supervivencia celular.

Palabras clave: Centralidad, Red reguladora, interacciones génicas, Factores de transcripción, Ley de potencia.

1 Objetivos

En esta segunda entrega se busca efectuar:

- Generar medidas de centralidad en la red génica reguladora de *Saccharomyces cerevisiae* como método para identificar agentes relevantes dentro de la red dirigida y para ello proponer una medida de *Esencialidad* de los genes.
- Comparar las medidas de centralidad y sus combinaciones respecto a los roles independientes de cada gen en la red génica.
- Determinar si existe una ley de potencia asociada que brinde información respecto a la estructura y tipo de red.

- Caracterizar el comportamiento de distribuciones de frecuencia a partir de las distintas medidas de centralidad.

2 Introduccción

En la red reguladora de *Saccharomyces cerevisiae* iMH805/775, la estructura jerárquica que involucra factores de transcripción (TFs), señales reguladoras y la expresión génica ofrece una visión detallada de cómo las células responden a cambios en el entorno, como la disponibilidad de nutrientes. Esta red funciona de manera que las señales reguladoras —generalmente relacionadas con factores nutricionales o ambientales— desencadenan

respuestas específicas dentro de la célula, modulando la actividad de los Regulator TFs, que son los principales actores en este proceso de regulación. Estos Regulator TFs controlan directamente la actividad de otros Target TFs, que a su vez regulan la transcripción de genes diana involucrados en procesos metabólicos y de crecimiento [1].

La señal reguladora actúa en lo más alto de la jerarquía, como una especie de interruptor que activa o reprime la actividad de los Regulator TFs dependiendo de las condiciones externas. Por ejemplo, en presencia de ciertos nutrientes, una señal puede activar un Regulator TF, lo que resulta en la activación de genes que promueven el metabolismo necesario para procesar esos nutrientes. En otros casos, la señal podría desencadenar la represión de la actividad de los Target TFs, disminuyendo la transcripción de genes que no son necesarios bajo ciertas condiciones.

Los Target TFs, por su parte, son los receptores finales de esta cascada de control. Su papel es crucial porque son los que realmente determinan si un grupo de genes específicos se transcribe o no. En esta red, el proceso es altamente controlado y permite a la célula ajustar con precisión sus respuestas metabólicas y de desarrollo en función de los recursos disponibles. Por tanto, los Regulator TFs tienen un papel destacado en esta red, ya que pueden tener múltiples interacciones, regulando varios Target TFs, lo que los convierte en nodos clave dentro de la red [2].

2.1 Métodos propuestos para identificar genes esenciales

- Uso de índices de centralidad clásicos individuales: este método implica la aplicación de medidas de centralidad tradicionales, como las centralidades de grado, cercanía e intermediación, a las redes biológicas. Estas medidas ayudan a identificar nodos altamente conectados, que a menudo son esenciales.
- Combinación de medidas de centralidad clásicas: los investigadores han descubierto que ninguna medida de centralidad puede predecir de manera confiable los genes esenciales. Por lo tanto, la combinación de múltiples índices de centralidad puede generar mejores predicciones. Por ejemplo, los estudios han demostrado que la combinación de coeficientes de agrupamiento con otras medidas, como la cercanía o la radialidad, mejora la identificación de genes

esenciales.

2.2 Medidas de centralidad

- Centralidad de proximidad: esta medida evalúa qué tan cerca está un nodo de todos los demás nodos de la red. De (referencia) destacan que identifica una fracción significativa de genes esenciales cuando se aplica a redes específicas. Es descrita como sigue por el inverso de la distancia media:

$$C_u = \frac{n-1}{\sum_{w \in V} \text{dist}(u, w)} \quad (1)$$

- Centralidad de intermediación: esta medida evalúa la cantidad de veces que un nodo actúa como un puente a lo largo del camino más corto entre otros dos nodos. A menudo se utiliza para identificar nodos clave que controlan la comunicación dentro de la red. Representa la fracción de caminos más cortos dentro de la red, que utilizan el nodo u , es descrita por:

$$C_u = \sum_{s \neq t \neq u \in V} \frac{\sigma_G(s, t|u)}{\sigma_G(s, t)} \quad (2)$$

- Centralidad de grado: esta medida cuenta la cantidad de conexiones directas que tiene un nodo. Puede indicar la importancia de un gen en función de sus interacciones directas con otros genes. Con esta medida se obtiene el número de conexiones al nodo u /salientes desde el nodo u .

$$C_u = k_{in}(u)/k_{out}(u) \quad (3)$$

- Centralidad de Katz: esta medida considera la cantidad de caminos que conducen a un nodo, lo que da más peso a los caminos más cortos. Refleja la influencia de un nodo en función de sus conexiones. Indica que un nodo tiene un valor c_{katz} mayor mientras más rutas lleguen a él, se determina como:

$$C_{katz} = \sum_{a=1}^{\infty} \alpha^a (A^T)^a \mathbf{1} \quad (4)$$

- PageRank: Esta medida evalúa la importancia de un nodo en función de la calidad y cantidad de sus conexiones, de forma similar a cómo Google clasifica las páginas web. Es útil para identificar nodos

influyentes en una red. La centralidad de un nodo depende de sus conexiones entrantes y de la conectividad relativa de estas conexiones, esto se calcula por medio de:

$$c_{PR} = dPc_{PR} + ((I - d)I) \quad (5)$$

- Coeficiente de clusterización: es una medida que cuantifica la tendencia de los nodos en una red a formar grupos o clústeres. En el contexto de redes biológicas, un alto coeficiente de clusterización indica que los genes (nodos) están más interconectados entre sí, lo que sugiere que pueden estar involucrados en funciones biológicas similares o en rutas metabólicas comunes. Este coeficiente se utiliza para identificar genes que tienen conexiones fuertes con sus vecinos, lo que puede ser indicativo de su importancia en la red. Para determinar este coeficiente se usa la fórmula:

$$C_u = \frac{2E_u}{k_u(k_u - 1)} \quad (6)$$

donde E_u es el número de conexiones (aristas) entre los vecinos del nodo u y k_u es el grado del nodo u .

Esta fórmula mide la proporción de conexiones existentes entre los vecinos de un nodo en comparación con el número total de conexiones posibles entre esos vecinos. Un coeficiente de clusterización cercano a 1 indica que los vecinos de un nodo están altamente interconectados, mientras que un valor cercano a 0 indica que no lo están. La **Centralidad conjunta** está dada por la siguiente fórmula:

$$C_S = \frac{1}{2} \sum_{i=1}^m \left(\frac{C(v_i) - C_{\min}}{C_{\max} - C_{\min}} \right)^2 \quad (7)$$

donde:

- C_S es la centralidad conjunta.
- $C(v_i)$ es la centralidad del nodo v_i .
- C_{\max} es el valor máximo de la centralidad $C(v)$.
- C_{\min} es el valor mínimo de la centralidad $C(v)$.
- m es el número total de nodos.

2.3 Breve estado del arte

Las medidas de centralidad no solo han sido usadas en redes dirigidas como [1], sino en redes de coexpresión génica como [3] donde los links representan la probabilidad coexpresión de dos genes. No obstante, las medidas de centralidad clásicas están diseñadas para sistemas específicos, normalmente sistemas sociales, por esta razón, se han diseñado nuevos métodos de centralidad basados en motifs [4], patrones locales funcionales que tienen las redes biológicas.

No obstante, parte de la importancia de estas redes está en su dinámica, estudiada por diferentes modelos como el de ising en una red dirigida tipo génica [5] donde se pueden determinar fases, o como métodos globales y estadísticos que estudian propiedad de la red en su totalidad [6].

3 Metodología

La implementación computacional de medidas de centralidad en redes genéticas involucra varios pasos y metodologías clave. A continuación se presentan los principales componentes de la implementación computacional:

3.1 Adquisición de datos y construcción de redes

- **Uso de bases de datos:** Las redes genéticas reguladoras se construyen utilizando bases de datos como Yeast - *Saccharomyces cerevisiae* y modelos curados como iMH805/775 [7] a los cuales se les hizo una limpieza previa en pro de relacionar las funcionalidades entre los elementos involucrados. Cabe resaltar que estas bases de datos proporcionan información sobre interacciones genéticas, rutas metabólicas y funciones enzimáticas que no serán tenidas en cuenta en su totalidad para este estudio.
- **Representación de la red:** Las redes se representan como grafos, donde los nodos corresponden a genes (o factores de transcripción) y los bordes representan interacciones o relaciones entre ellos, en este caso acciones de regulación para activar o suprimir tanto los genes o los factores de transcripción a través de señales y otros TF principales.

3.2 Cálculo de medidas de centralidad

- **Selección de medidas de centralidad:**

Se seleccionan varias medidas de centralidad para el análisis bajo recomendaciones de [1], incluyendo cercanía, intermediación, grado, Katz y PageRank, así como un promedio de algunas de ellas. Cada medida cuantifica la importancia de los nodos de diferentes maneras, según su posición y conexiones dentro de la red, todo esto a través del paquete de Python **Networkx** y sus distintas funciones .

- **Implementación de algoritmos:** Los algoritmos para calcular estas medidas de centralidad se implementan utilizando la librería **Networkx** en Python que facilita el cálculo eficiente de estas métricas.

4 Resultados

Cuando se grafican estos nodos y conexiones como se ve en la figura 1 podemos observar visualmente cómo se estructuran estas relaciones de activación o represión, además de como se distribuyen los distintos tipos de nodos dependiendo si son señales, factores de transcripción reguladores o genes objetivo. La figura 2 muestra el mapa de calor según el grado total de cada nodo. Se observa que los nodos más al centro del grafo son los de mayor grado, comparando con 1 cuyas posiciones de los nodos son las mismas, se puede observar que las señales y los reguladores son los factores más conectados, mientras que los genes objetivo, que luego se expresan son los menos conectados.

Los nodos "no activados o reprimidos" representan aquellos elementos que no están siendo influidos por ninguna señal reguladora en un momento dado. Los enlaces influidos de manera "activadora" y "represora" muestran la dinámica de la red, con colores que reflejan si los genes asociados están siendo activados (aumentando la expresión génica) o reprimidos (disminuyendo dicha expresión). El grafo obtenido se muestra en la siguiente figura:

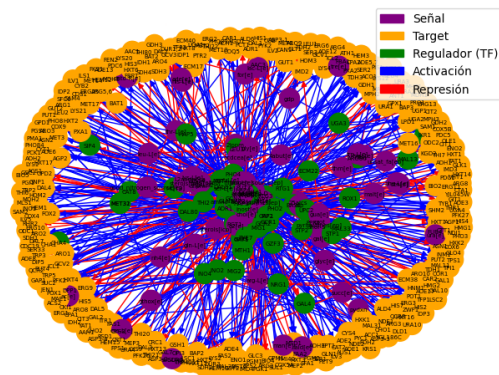


Figure 1: Grafo dirigido para la red metabólica de la *Saccharomyces cerevisiae* iMH805/77, clasificado por tipo de nodo.

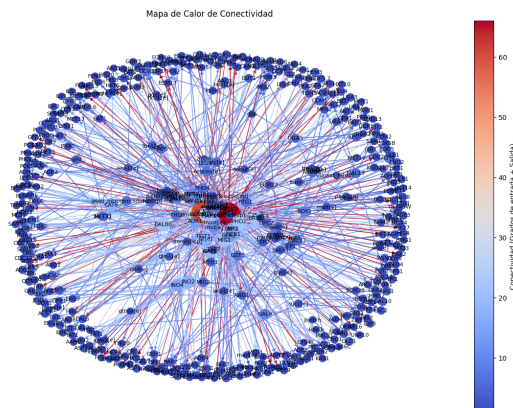


Figure 2: Mapa de Calor de la conectividad la red metabólica de la *Saccharomyces cerevisiae* iMH805/77.

4.1 Ley de Potencias en la Red

Este primer resultado da cabida a analizar otras bases de datos asociados para la *Saccharomyces cerevisiae*, según sea una red regulatoria o metabólica pues estas dependen de funcionalidades internas particulares y distintas para cada caso. También se determinó una distribución de la frecuencia de los grados en la red para ver si tenía asociada una ley de potencia (reflejando una organización jerárquica donde algunos nodos regulan un gran número de conexiones, lo cual es común en redes biológicas) de la forma $P(k) \propto k^{-\alpha}$, de la cuál se determinó un exponente $\alpha = 2.6347$, así mismo se hizo un análisis de la frecuencia en función de los valores de centralidad para las distintas medidas realizadas sobre la red, dando información tanto de la estructura como del tipo de red sugiriendo que es una red

libre de escala, caracterizada por la presencia de nodos clave o hubs que tienen un número elevado de conexiones, mientras que la mayoría de los nodos tienen pocos enlaces. Del mismo modo notamos que la estructura topológica corresponde a redes de Barabási-Albert al tener una jerarquía asociada de **Señales** \rightarrow **Factores de transcripción** \rightarrow **Genes objetivo** que se destacan en la red por sub-categorías con agentes principales en función de su grado-conectividad, como se verá más adelante. Dicha estructura implica que la red es robusta frente a la eliminación aleatoria de nodos, aunque la pérdida de un hub crucial podría fragmentar la red. Dicho ajuste y suavizado se muestra a continuación:

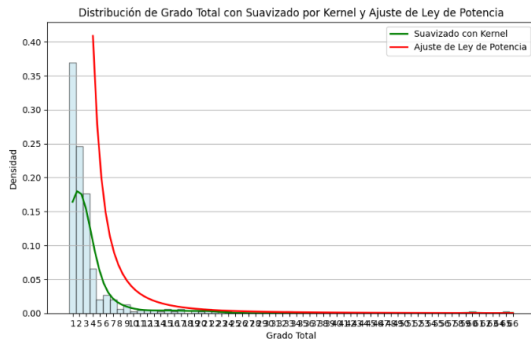


Figure 3: Ajuste de Ley de potencia y suavizado de la distribución de grado total.

Al aplicar medidas de centralidad sobre este grafo dirigido (las antes mencionadas implementando NetworkX en Python), para aproximarnos a la identificación de los genes esenciales se procedió comparando a través de dos medidas que dieran información de la relevancia individual para así comparar y destacar en un "top" cuáles fueron esos elementos, en el primer caso de las 6 medidas de centralidad lo que se hizo fue promediarlas para determinar un puntaje de "centralidad común", en el otro caso y siguiendo el ejemplo de la referencia [1] determinamos un puntaje dado por la Centralidad conjunta de la ecuación (7) a partir de los otros resultados obtenidos, esto para cada gen dando cuenta que entre más pequeño sea este puntaje mayores serán sus medidas de centralidad, entre estas dos medidas se pueden identificar elementos comunes, a través del recurso [7] se pudo identificar las funcionalidades particulares siendo estas relevantes para el desempeño de la red génica, sin embargo estas medidas no son suficientes pues es necesario tener más en cuenta la dinámica interna del sistema, sea por asignación de pesos en los enlaces tal que estos cambian la estructura por la influencia que

tienen sobre los genes objetivos, razón que daría más utilidad a la centralidad combinada permitiendo filtrar los genes más relevantes en las funciones internas de la red.

4.2 Análisis de Centralidades

A continuación se presentan las distribuciones de los valores para cada centralidad con conclusiones complementarias asociadas respecto a la función que desempeñan los elementos rankeados en cada gráfica y tabla.

Centralidad Out

La centralidad de grado out es quizá una de las más reelevantes pues presentan valores más pronunciados, por ejemplo, los primeros dos valores son de 66 y 57 (los rojos de la figura 2) seguidos de valores entre 30 y 20. Estos dos corresponden a la señal *glc[e]* y al factor *GCN4*, la primera encargada del transporte de glucosa que media su consumo de esta del ambiente, mientras que la segunda está asociada a la producción de aminoácidos en escasez.

En general, estos genes que se muestran en la tabla 1 tienen un gran impacto en la célula de *Saccharomyces cerevisiae* porque controlan procesos clave en el metabolismo y la respuesta a cambios en el ambiente, como la disponibilidad de nutrientes.

Algunos de estos factores regula un conjunto de genes que permiten a la célula adaptarse a diferentes situaciones, como la falta de glucosa o aminoácidos. De este modo, esta medida de centralidad puede señalar factores importantes que regulan procesos esenciales, tales como el metabolismo de glucosa, nitrógeno y fosfato, que influyen en muchas otras funciones de la célula. En este sentido, muestra algunos "interruptores maestros" que aseguran que la célula esté funcionando de manera eficiente bajo distintas condiciones.

Table 1: Top 10 Nodos con mayor Centralidad de Grado de Salida

Nodo	G. Salida	Tipo de Nodo
<i>glc[e]</i>	66	signal
<i>GCN4</i>	57	TF
<i>GLN3</i>	31	TF
<i>HAP2</i>	23	TF
<i>PHO2</i>	21	TF
<i>HAP3</i>	20	TF
<i>MIG1</i>	20	TF
<i>DAL80</i>	18	TF
<i>BAS1</i>	18	TF
<i>o2[e]</i>	17	signal

La mayoría de los nodos tienen valores de centralidad de salida cercanos a cero. La distribución altamente asimétrica muestra que la mayoría de los nodos son relativamente menos influyentes en términos de regulación directa, mientras que unos pocos tienen una gran capacidad de control sobre otros nodos.

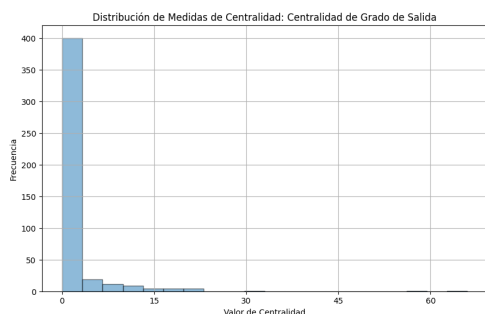


Figure 4: Centralidad de Grado de salida.

Centralidad In

Por otro lado, la centralidad de entrada (In) no muestra mucha variación, llegando a un máximo de grado 7 y siendo el top 10 nodos target (genes objetivo).

Estos genes están altamente regulados porque sus funciones son críticas para que la célula se adapte rápidamente a cambios en su entorno, como la falta de nutrientes o la necesidad de producir más energía. La célula usa muchos factores de control para decidir cuándo activar o desactivar estos genes según sus necesidades.

En la tabla de abajo se puede ver que estos son en su mayoría genes que actúan como objetivos (targets) de factores de transcripción (FT). Esto indica que reciben múltiples interacciones regulatorias. Entre ellos, CIT1 y GCV3 son los nodos más conectados, recibiendo 7 conexiones. Esto sugiere que están involucrados en procesos metabólicos importantes, como la respiración mitocondrial (CIT1) y el metabolismo de glicina (GCV3). Otros genes como SUC2, JEN1, y GDH1 también tienen un alto grado de entrada, lo que indica su importancia en el metabolismo de azúcares y aminoácidos.

Table 2: Top 10 Nodos con mayor Centralidad de Grado de Entrada

Nodo	G. Entrada	Tipo de Nodo
CIT1	7	target
GCV3	7	target
SUC2	6	target
JEN1	6	target
GDH1	6	target
UGA4	6	target
CYB2	6	target
GLN1	5	target
LPD1	5	target
GCV2	5	target

La distribución de la centralidad de entrada tiene una forma similar a una distribución de Poisson, donde la mayoría de los nodos tienen grados de entrada bajos (cerca de 1 o 2). Esto sugiere que la mayoría de los genes objetivo en la red son regulados por un número limitado de factores de transcripción o señales, lo que refleja un control distribuido de la red. A medida que el valor de la centralidad aumenta, la frecuencia disminuye rápidamente, y solo unos pocos nodos tienen valores de grado de entrada altos (como CIT1 y GCV3). Esto podría estar relacionado con la especialización de ciertas funciones metabólicas altamente reguladas.

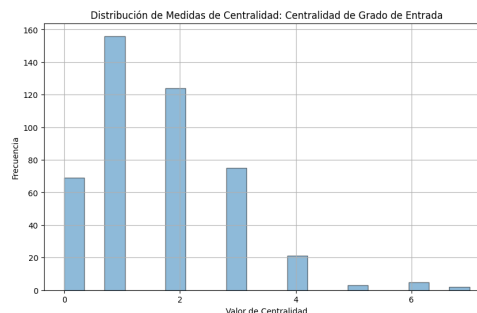


Figure 5: Centralidad de Grado de Entrada.

Centralidad de Intermediación

La centralidad de intermediación por otro lado considera ambos grados, de entrada y de salida, midiendo la cantidad de caminos más cortos que intermedia un nodo dado. Los nodos más intermediarios se muestran en la tabla 3. Todos son factores de transcripción puesto que estos son los elementos intermedios en la red considerada, además, aparecen varios nodos que de por sí ya tenían grado alto. Sin embargo aparecen nuevos nodos como los factores CAT8 y HAP4, estos tienen una particularidad y es que ambos son partícipes

de la regulación de la respiración celular, en este proceso también están los factores MIG2 y MIG1. Esto podría ser mera casualidad o significar que hay procesos cuyos nodos tienen una alta intermediación.

Esta métrica resalta los nodos que actúan como puentes entre diferentes partes de la red GCN4 tiene la mayor intermediación, lo que refuerza su papel crítico en conectar múltiples rutas regulatorias. Otros nodos como CAT8 y HAP4 son también intermediarios importantes, relacionados con la regulación del metabolismo de carbohidratos.

Table 3: Top 10 Nodos con mayor Centralidad de Intermediación

Nodo	C. Intermediación	Tipo de Nodo
GCN4	0.000807	TF
CAT8	0.000608	TF
HAP4	0.000462	TF
MIG2	0.000336	TF
GLN3	0.000272	TF
BAS1	0.000238	TF
RGT1	0.000209	TF
MIG1	0.000132	TF
GAT1	0.000131	TF
DAL80	0.000119	TF

La distribución de centralidad de intermediación está fuertemente sesgada hacia valores extremadamente bajos, con unos pocos nodos mostrando centralidades mucho más altas. Esto indica que la mayoría de los nodos en la red no actúan como intermediarios importantes en los caminos más cortos de la red, mientras que un número muy reducido de nodos (como GCN4 y CAT8) juegan un papel crítico en conectar diferentes partes de la red. Estos nodos con alta intermediación actúan como "puentes" que facilitan la comunicación y el flujo de información a través de la red. La baja frecuencia de nodos con valores altos refuerza la idea de que solo algunos factores tienen roles esenciales en la regulación de procesos.

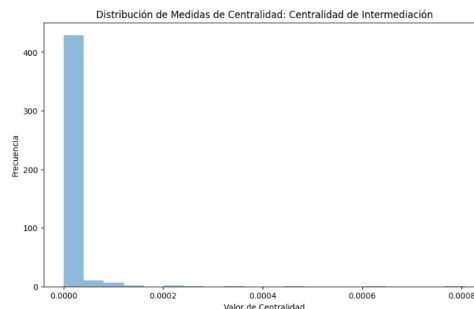


Figure 6: Centralidad de Intermediación.

Es importante resaltar que los únicos nodos que pueden intermediar en este tipo de red son los TFs, esto ya que la red ignora otras relaciones post-transcripción, por esta razón esta medida es especial para determinar cierto grado de esencialidad en el fenómeno que se analice, que en este caso, es la regulación de genes. En la tabla

Aquí se destacan nodos target con alta accesibilidad a otros nodos en la red: GDH1, GLN1, y CIT1 están entre los más prominentes, lo que sugiere que son altamente accesibles por los factores de transcripción y están involucrados en rutas clave del metabolismo de aminoácidos y la respiración.

Proximidades In y Out

Table 4: Top 10 Nodos con mayor Centralidad de Proximidad In

Nodo	Proximidad In	Tipo de Nodo
GDH1	0.020389	target
GLN1	0.019824	target
CIT1	0.018722	target
GCV3	0.016943	target
DAL1	0.016920	target
DUR1,2	0.015859	target
DAL7	0.015859	target
DAL4	0.015859	target
DAL2	0.015859	target
DUR3	0.015859	target

Se muestran nodos con acceso eficiente a otros nodos como glc[e] tiene la mayor proximidad de salida, lo que refleja cómo la glucosa extracelular puede rápidamente influir en otros procesos a través de múltiples factores de transcripción. Los FT GCN4 y GLN3 también tienen alta proximidad de salida, mostrando su capacidad para influir rápidamente en diversas rutas metabólicas.

Table 5: Top 10 Nodos con mayor Centralidad de Proximidad Out

Nodo	Proximidad Out	T. Nodo
glc[e]	0.216930	signal
GCN4	0.125551	TF
amino_acids[ep]	0.088350	signal
purines[ep]	0.074907	signal
o2[e]	0.072710	signal
nh4[e]	0.067774	signal
GLN3	0.066630	TF
MIG1	0.064997	TF
HAP2	0.050661	TF
PHO2	0.046256	TF

En el caso de ambas proximidades, los nodos de señalización y los factores de transcripción, con alta proximidad out, juegan un papel crucial en la regulación de respuestas rápidas y efectivas, lo que los posiciona como reguladores clave.

Los nodos que reciben muchas conexiones (nodos con alto grado de entrada) también son centrales en términos de proximidad, lo cual tiene sentido, ya que estos nodos probablemente actúan como puntos de convergencia en la red regulatoria. Estos nodos podrían ser cruciales para la coordinación de procesos globales en la célula, dado que son genes target influidos cercanamente por distintos factores reguladores, estos pueden ser encargados de una respuesta rápida a los cambios en las necesidades de la célula.

A pesar de esto, la red solo considera la topología de las interacciones mas no su velocidad, por esta razón, es necesario considerar pesos para estas redes.

La centralidad de Katz destaca la relevancia de nodos que pueden estar conectados indirectamente a factores clave o señales metabólicas, lo que implica que estos genes están en posiciones estratégicas para influir en varios procesos celulares. Por ejemplo, CIT1, que tiene el valor más alto, es fundamental en el ciclo de Krebs, un proceso central en la generación de energía celular. De manera similar, UGA4 y GCV3 están implicados en el metabolismo de fuentes de nitrógeno y la biosíntesis de aminoácidos, respectivamente, lo que resalta su rol en la respuesta adaptativa a cambios en el ambiente. A continuación se muestra el listado obtenido:

Table 6: Top 10 Nodos con mayor Centralidad de Katz

Nodo	C. de Katz	T. de Nodo
CIT1	0.070109	target
UGA4	0.068428	target
GCV3	0.067638	target
GDH1	0.066551	target
JEN1	0.065504	target
CYB2	0.064596	target
SUC2	0.064592	target
GLN1	0.063586	target
GCV2	0.063346	target
LPD1	0.061078	target

De la figura 7 se destacan los nodos que no solo tienen muchas conexiones directas, sino que también están conectados indirectamente a nodos influyentes. Esto implica que algunos nodos con grado de entrada alto también son fundamentales en el contexto global de la red, ya que incluso los nodos más distantes dependen de ellos.

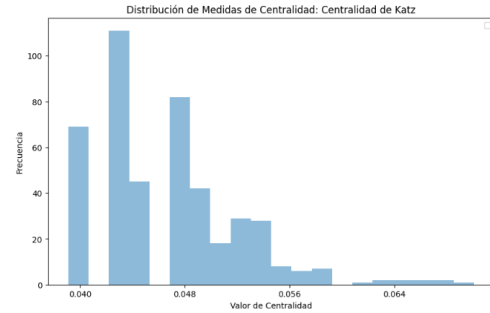


Figure 7: Centralidad de Katz.

Aquí PPR1, un factor de transcripción, tiene el PageRank más alto, lo que sugiere su importancia en la regulación de múltiples genes. GCV3 también aparece como un target importante en esta medida, relacionado con el metabolismo del nitrógeno y la glicina.

Table 7: Top 10 Nodos con mayor PageRank

Nodo	PageRank	Tipo de Nodo
PPR1	0.006898	TF
GCV3	0.006626	target
BAS1	0.005423	TF
ARO80	0.005423	TF
CYS3	0.004809	target
PUT3	0.004686	TF
DAL82	0.004686	TF
LYS14	0.004686	TF
ARO10	0.004040	target
ARO9	0.004040	target

Se puede resaltar que hay pocos nodos con un valor de PageRank alto y muchos nodos con valores bajos. Esto indica que la red tiene una estructura jerárquica en la que unos pocos nodos son especialmente importantes en términos de la acumulación de influencia a través de sus conexiones, mientras que la mayoría de los nodos son menos influidos.

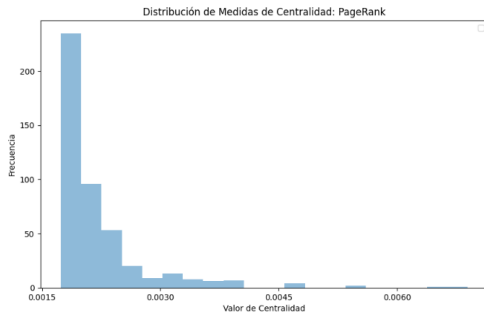


Figure 8: Centralidad de PageRank.

En la figura de abajo se puede ver que la clusterización es por lo general baja, y pocos nodos comparten vecinos entre sí. De este modo, es posible que las rutas regulatorias que son redundantes o que cumplen la misma función para condiciones distintas sean más diversas que los triángulos.

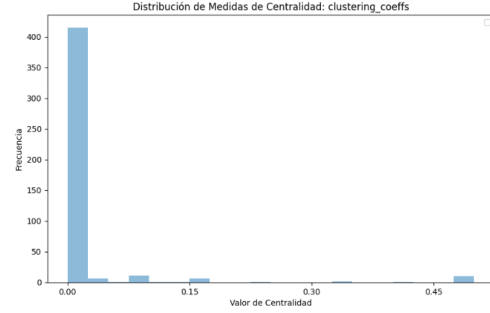


Figure 9: Coeficiente de Clusterización.

Los nodos que actúan como intermediarios y tienen un alto grado de salida (como la señal glc[e] y el factor GCN4) parecen tener menor tendencia a estar en clústeres muy cohesivos, lo que refleja su papel como conectores entre distintas regiones de la red. Estos nodos son más responsables de integrar información que de pertenecer a un grupo muy interconectado, lo cual tiene sentido desde el punto de vista funcional. La similitud con el grado de salida también sugiere que los nodos que regulan muchos otros genes no forman parte de comunidades densamente conectadas. Estos nodos probablemente actúan de manera más global, controlando diferentes funciones en la célula a través de conexiones distribuidas por toda la red.

Para los nodos UGA4, GCV2, y los genes de la familia ASP3 son los más destacados. Estos nodos están asociados con la absorción y utilización de nitrógeno, lo que refuerza su rol central en el metabolismo de este elemento. La presencia de los factores de transcripción DAL80 y GAT1 también sugiere que, aunque los genes objetivo son prominentes en esta métrica, los reguladores clave no están ausentes.

Table 8: Top 10 Nodos con mayor Centralidad de Vector Propio

Nodo	Vector Propio	Tipo de Nodo
UGA4	0.342110	target
GCV2	0.342110	target
ASP3-2	0.261349	target
ASP3-3	0.261349	target
ASP3-4	0.261349	target
GAP1	0.261349	target
ASP3-1	0.261349	target
DAL80	0.211436	TF
GAT1	0.211436	TF
ECM38	0.211436	target

La siguiente tabla resalta que las señales externas,

como la good nitrogen source[ep] y la señal glc[e], junto con factores de transcripción como GZF3, GAT1, y GLN3, son los elementos más influyentes.

Table 9: Top 10 Nodos con mayor Centralidad de Vector Propio reverse

Nodo	V. P. reverse	Nodo
good_nitrogen_source[ep]	5.373297e-01	signal
GZF3	4.347088e-01	TF
GAT1	4.347088e-01	TF
GLN3	4.347088e-01	TF
glc[e]	2.686648e-01	signal
DAL80	2.686648e-01	TF
MIG1	8.739349e-17	TF
MTH1	8.686907e-17	TF
malt[e]	5.688683e-17	signal
MIG2	5.467806e-17	TF

Por último, de la medida propuesta tanto glc[e] y GCN4 tienen los valores más altos, confirmando que tanto las señales externas como los factores de transcripción son cruciales para el control global del metabolismo.

Table 10: Top 5 Nodos por Promedio de Centralidades y Grado Total

Nodo	Prom. C.	Grado Total	Ranking
glc[e]	16.543869	66	41.271934
GCN4	15.015578	60	37.507789
GLN3	8.296245	33	20.648122
HAP2	6.008150	24	15.004075
PHO2	5.507836	22	13.753918

5 Conclusiones y Discusión

La estructura de la red es parecida a una invariante de escala con un exponente aproximado de -2.6 , lo que es señal de una autoorganización en el interior de la célula. Esto se asocia además con una jerarquía de influencia **Señales** \rightarrow **Factores de transcripción** \rightarrow **Genes objetivo**, donde los últimos presentan en general bajo grado a diferencia de las señales y los FTs que presentan alta conectividad.

Cada medida de centralidad puede dar resultados interesantes según su definición y aquello que la motiva. La centralidad de grado de salida tiene un especial interés puesto que los nodos más conectados presentan una clara diferencia con el resto, mostrando factores muy influyentes en la red. La centralidad de intermediación también es interesante puesto que

dice como influye un factor de transcripción en las rutas regulatorias, no obstante, es necesario agregarle pesos a los vínculos como el tiempo en que los factores se activan o desactivan entre sí. Por otra parte, la centralidad de proximidad puede indicar los factores de transcripción, como señales, que tienen una respuesta más rápida de la célula, o genes que así mismo respondan más rápido a las señales, encargados de la adaptabilidad de la célula.

Por otra parte, dado que la esencialidad y otras propiedades de factores y de genes depende de las interacciones dentro de la red, sería interesante indagar por resultados más precisos a través de nuevos conceptos de centralidad que han desarrollado nuevos índices y que incorporan características adicionales asociadas con los nodos, como la conectividad y los roles funcionales, para predecir mejor la esencialidad. Así mismo, la integración de datos ómicos con medidas de centralidad, este enfoque llamado centralidad basada en motifs [4], combina el conocimiento biológico con roles topológicos, mejorando el poder predictivo de las medidas de centralidad.

Se podría estudiar la información dinámica asociada a la transmisión de información respecto a las funciones que desempeñan los distintos agentes, en caso de ser posible la inclusión de tiempos de paso medio para un caminante aleatorio dentro de la red dirigida, además de un ranking como propuesta complementaria a las otras medidas de **esencialidad** previas, todo esto a partir de la matriz de transición/adyacencia del grafo dirigido tal y como lo efectúa en "Random Walks on Complex Networks".

References

- [1] Coello Del Rio Koschützki. "How to Identify Essential Genes from Molecular Networks?" *in BMC Systems Biology*: (2009).
- [2] Gupta Mahdi Salehzadeh-Yazdi. "Evolution of Centrality Measurements for the Detection of Essential Proteins in Biological Networks". *in Frontiers in Genetics*: (2016).
- [3] Francisco J Azuaje. "Selecting biologically informative genes in co-expression networks with a centrality score". *in Biology direct*: 9 (2014), pages 1–23.
- [4] Dirk Koschützki, Henning Schwöbbermeyer and Falk Schreiber. "Ranking of network elements based on functional substructures". *in Journal of theoretical biology*: 248.3 (2007), pages 471–479.

- [5] Jeyashree Krishnan **and others**. “A modified Ising model of Barabási–Albert network with gene-type spins”. **in** *Journal of mathematical biology*: 81 (2020), **pages** 769–798.
- [6] Lesley T MacNeil **and** Albertha JM Walhout. “Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression”. **in** *Genome research*: 21.5 (2011), **pages** 645–657.
- [7] Saccharomyces Genome Database. *YeastGenome.org*. <https://www.yeastgenome.org>.